Lecture 8-b
Time Series:
Stationarity, AR(p) & MA(q)

Brooks (4th edition): Chapter 6

 1

Review: Times Series

• A time series y_t is a process observed in sequence over time, t = 1, ..., T $\Rightarrow$ $Y_t = \{y_1, y_2, y_3, ..., y_T\}$.

• Given the sequential nature of y_t, we expect y_t & y_{t-1} to be dependent This is the main feature of time series: *dependence*.

• With dependent observations, the classical results (based on LLN & CLT) are not to valid. New assumptions and tools are needed: stationarity, ergodicity, & CLT for martingale difference sequences.

• Roughly speaking, stationarity requires constant moments for y_t; ergodicity requires that the dependence is short-lived, eventually y_t has only a small influence on y_{t+k}, when k is relatively large.

Review: Times Series – Forecasting & WN

• The purpose of building a time series model: Forecasting.

• We estimate time series models to forecast out-of-sample. For example, the *l-step ahead* forecast: $\hat{y}_{T+l} = \mathrm{E}_t[y_{t+l}\,|I_t]$, where $I_t = \{y_{t-1}, y_{t-2}, y_{t-3}, ...\}$

• Two popular models for $\mathrm{E}_t[y_t|I_t]$:

– Autoregressive (AR) process models $\mathrm{E}_t[y_t|I_t]$ with lagged y_t's:

$$\mathrm{E}_t[y_t|I_t] = f(y_{t-1}, y_{t-2}, y_{t-3}, ... , y_{t-p})$$

Example: AR(1) process, $y_t = \alpha + \beta\, y_{t-1} + \varepsilon_t$.

– Moving average (MA) process models $\mathrm{E}[y_t|I_{t-1}]$ with lagged ε_t's:

$$\mathrm{E}_t[y_t|I_t] = f(\varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, ... , \varepsilon_{t-q})$$

Example: MA(1) process, $y_t = \mu + \theta_1\, \varepsilon_{t-1} + \varepsilon_t$

Review: Times Series – Forecasting (again)

• We want to select an appropriate time series model to forecast y_t. The linear models we consider: AR(p), MA(q) or ARMA(p, q).

• Steps for forecasting:

(1) Identify the appropriate model. That is, determine AR, MA or ARMA and the order of the model -i.e., p, q.

Tools: ACF, PACF

(2) Estimate the model.

OLS, Method of Moments (complicated).

(3) Test the model.

Make sure errors are WN.

(4) Forecast.

Review: Times Series – Conditionality

• Key distinction: **Conditional** vs **Unconditional moments**.

Example: AR(1) process: $y_t = \alpha + \beta\, y_{t-1} + \varepsilon_t$.

The **conditional mean** forecast at time *t*, conditioning on I_{t-1}, is:

$$\mathrm{E}_t[y_t|I_t] = \mathrm{E}_t[y_t] = \alpha + \beta\, y_{t-1}$$

The **unconditional mean** is given by:

$$\mathrm{E}[y_t] = \alpha + \beta\, \mathrm{E}[y_{t-1}] = \frac{\alpha}{1-\beta} = \mu = \text{constant} \qquad (\beta \neq 1)$$

The conditional mean is time varying; the unconditional mean is not!

Remark: Time series focuses on conditional forecasts.

Review: Stationarity

• We say that a process is stationary of

1st order if $F(y_{t_1}) = F(y_{t_{1+k}})$ for any t_1, k

2nd order if $F(y_{t_1}, y_{t_2}) = F(y_{t_{1+k}}, y_{t_{2+k}})$ for any t_1, t_2, k

Nth-order if $F(y_{t_1}, \ldots, y_{t_T}) = F(y_{t_{1+k}}, \ldots, y_{t_{T+k}})$ for any t_1, ..., t_T, k

• We focus on ***2nd order* stationarity**, which is weaker: only consider mean and covariance (& easier to verify in practice). Thus, we need

$$\mathrm{E}[Y_t] = \mu$$
$$\mathrm{Var}(Y_t) = \sigma^2 = E[(Y_t - \mu)^2]$$
$$\mathrm{Cov}(Y_{t_1}, Y_{t_2}) = E[(Y_{t_1} - \mu)(Y_{t_2} - \mu)] = \gamma(t_1 - t_2) = \gamma(k)$$

Notes: $\gamma(k)$: **autocovariance function**, a function of $k = t_1 - t_2$.
$\gamma(0)$ is the variance.

Review: Stationarity & Autocovariances

- From the autocovariances, we derive the autocorrelations:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \rho(k)$$

$\rho(k)$, a function of k, is called the **auto-correlation function** (**ACF**).

- The ACF is one of the tools used to identify a model: MA(q), AR(p).

Review: Stationarity – Example

Example: Assume $\varepsilon_t \sim$ WN(0, σ^2).

$y_t = \phi\, y_{t-1} + \varepsilon_t$. (AR(1) process)

- **Mean**

  E[y_t] = μ = 0 (assuming $\phi \neq 1$)

- **Variance**

  Var[y_t] = $\gamma(0) = \phi^2$ Var[y_{t-1}] + Var[ε_t]

  $\gamma(0) = \sigma^2/(1 - \phi^2)$ (assuming $|\phi| < 1$)

- **Covariance**

  $\gamma(k)$ = Cov[y_t, y_{t-k}] = $\phi^k\,\gamma(0)$

$\Rightarrow$ If $|\phi| < 1$, AR(1) process is covariance **stationary**.

- Auto-correlation function (**ACF**): $\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \phi^k$

Review: Non-Stationarity – Example

Example: Assume $\varepsilon_t \sim$ WN(0, σ^2).

$y_t = \mu + y_{t-1} + \varepsilon_t$ (Random Walk with drift process)

Doing backward substitution:

$\Rightarrow y_t = \mu\ t + \sum_{j=0}^{t-1} \varepsilon_{t-j} + y_0$

- **Mean** & **Variance**

  $E[y_t] = \mu\ t + y_0$

  $Var[y_t] = \gamma(0) = \sum_{j=0}^{t-1} \sigma^2 = \sigma^2\ t$

$\Rightarrow$ the RW process is **non-stationary**; that is, moments are time dependent.

Review: Stationarity: Remarks

• Stationarity is an invariant property: The statistical characteristics of the time series do not vary over time.

• If IBM is weak stationary, then, the returns of IBM may change month to month or year to year, but the average return and the variance in two equal lengths time intervals will be more or less the same.

• In the long run, say 100-200 years, the stationarity assumption may not be realistic.

• In general, time series analysis is done under the stationarity assumption.

Review: Ergodicity

• We want to estimate the mean of the process $\{Z_t\}$, $\mu(Z_t)$. But, we need to distinguishing between *ensemble average* (with m *cross section* observations) and *time average* (with T *time series* observations):

- Ensemble Average: $\bar{\bar{z}} = \frac{\sum_{i=1}^{m} Z_i}{m}$

- Time Series Average: $\bar{z} = \frac{\sum_{t=1}^{T} Z_t}{T}$

Q: Which estimator is the most appropriate?
A: Ensemble Average. But, it is impossible to calculate for a time series.

• The *Ergodic Theorem* tells us when the time series average can be used.

Theorem: A sufficient condition for ergodicity for the mean:

$$\rho_k \to 0 \quad \text{as} \quad k = t_i - t_j \to \infty$$

We need the correlation between (y_{t_i}, y_{t_j}) to decrease as they grow further apart in time.

Review: Invertibility

Example: Suppose we have an MA(1) process:

$$y_t = \mu + \theta_1 \varepsilon_{t-1} + \varepsilon_t = \mu + \theta(L)\varepsilon_t \qquad - \theta(L) = (1 + \theta_1 L)$$

Recall: $f(x) = \frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + \ldots = \sum_{n=0}^{\infty} x^n$

Let $x = -\theta_1 L$, (L: lag operator). Then, assuming $\theta(L) \neq 0$:

$$\theta(L)^{-1} = \sum_{n=0}^{\infty}(-\theta_1 L)^n = 1 - \theta_1 L + \theta_1^2 L^2 - \theta_1^3 L^3 + \theta_1{}^4 L^4 + \cdots$$

Now, we multiply $\theta(L)^{-1}$on both sides of the MA process

$$y_t = \mu + \theta(L)\, \varepsilon_t.$$

$$\Rightarrow \qquad \theta(L)^{-1}\, y_t = \theta(L)^{-1}\mu + \varepsilon_t = \mu^* + \varepsilon_t$$

Then, we get an AR(∞):

$$y_t = \mu* + \theta_1 y_{t-1} - \theta_1^2\, y_{t-2} + \theta_1^3\, y_{t-3} - \theta_1^4 y_{t-4} + \cdots + \varepsilon_t$$

If the AR(∞) process in non-explosive, then, the MA(1) is **invertible**.

Review: Moving Average Process

• An MA process models $E_t[y_t \mid I_{t-1}]$ with lagged error terms. An MA(q) model involves q lags.

• We keep ε_t ~ WN(0, σ^2)

Example: A linear MA(q) model:

$$y_t = \mu + \theta_1\, \varepsilon_{t-1} + \theta_2\, \varepsilon_{t-2} + \ldots + \theta_q\, \varepsilon_{t-q} + \varepsilon_t = \mu + \theta(L)\, \varepsilon_t,$$

where

$$\theta(L) = 1 + \theta_1\, L + \theta_2\, L^2 + \theta_2\, L^3 + \ldots + \theta_q\, L^q$$

• In time series, the constant does not affect the properties of AR and MA process. Thus, in this situation we say "without loss of generalization", we assume $\mu = 0$.

Review: Moving Average Process – Stationarity

• Q: Is MA(q) stationary? Check the moments (assume $\mu = 0$).

$$y_t = \varepsilon_t + \theta_1\, \varepsilon_{t-1} + \theta_2\, \varepsilon_{t-2} + \ldots + \theta_q\, \varepsilon_{t-q}$$

• **Mean**

$$E[y_t] = E[\varepsilon_t] + \theta_1\, E[\varepsilon_{t-1}] + \theta_2\, E[\varepsilon_{t-2}] + \ldots + \theta_q\, E[\varepsilon_{t-q}] = 0$$

• **Variance**

$$\text{Var}[y_t] = \text{Var}[\varepsilon_t] + \theta_1^2\, \text{Var}[\varepsilon_{t-1}] + \theta_2^2\, \text{Var}[\varepsilon_{t-2}] + \ldots + \theta_q^2\, \text{Var}[\varepsilon_{t-q}]$$
$$= (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2)\, \sigma^2.$$

• **Covariance**

In general, for the k *autocovariance:*

$$\gamma(k) = \sigma^2 \textstyle\sum_{j=k}^{q} \theta_j\, \theta_{j-k} \qquad \text{for } |\, k\,| \leq q$$
$$\gamma(k) = 0 \qquad \text{for } |\, k\,| > q$$

Remark: After lag q, the autocovariances (& ACFs) are 0.

Review: Moving Average Process – Stationarity

• It is easy to verify that the sums $\sum_{j=k}^{q} \theta_j \theta_{j-k}$ are finite. Then, mean, variance and covariance are constant.

⇒ **MA(q) is always stationary**.

• Problem: It can be shown that for ε_t with same distribution (say, normal) the autocovariances are non-unique. Suppose you want to select one process to forecast. Which one? We select the **invertible** model, with an AR(∞) (non-explosive) representation.

Technical note: An invertible MA(q) is typically required to have roots of the lag polynomial equation $\theta(z) = 0$ greater than one in absolute value (**outside the unit circle**). In the MA(1) case, we require $|\theta_1| < 1$

Review: MA Process – MA(1): Moments

Example: MA(1) process:

$$y_t = \theta_1 \varepsilon_{t-1} + \varepsilon_t = \mu + \theta(L)\, \varepsilon_t, \quad \text{with } \theta(L) = (1 + \theta_1 L)$$

• **Moments**

$E[y_t] = 0$

We derive the variance & autocovariances from the MA(q) formula:

$$\gamma(k) = \sigma^2 \sum_{j=k}^{q} \theta_j\, \theta_{j-k} \quad \text{for } |k| \leq q \qquad (\text{where } \theta_0 = 1)$$

$$\gamma(k) = 0 \quad \text{for } |k| > q$$

• $\gamma(k)$, with $q = 1$

$k = 0$ $\quad \gamma(0) = \sigma^2 \sum_{j=0}^{1} \theta_j\, \theta_{j-0} = \sigma^2 (1 + \theta_1^2)$

$k = 1$ $\quad \gamma(1) = \sigma^2 \sum_{j=1}^{1} \theta_j\, \theta_{j-1} = \sigma^2 (\theta_1)$

$k > 1$ $\quad \gamma(k) = 0$

Since the sums $\sum_{j=k}^{q} \theta_j\, \theta_{j-k}$ are finite ⇒ MA(q) is always stationary.

Review: MA Process – MA(1): ACF

• **Autocorrelations**

To get the ACF, we divide the autocovariances by $\gamma(0)$. Then:

$\rho(0) = \gamma(0)/\gamma(0) = 1$

$\rho(1) = \gamma(1)/\gamma(0) = \theta_1/(1+\theta_1^2)$

$\rho(2) = \gamma(2)/\gamma(0) = 0$

$\vdots$

$\rho(k) = \gamma(k)/\gamma(0) = 0$ (for $k > 1$)

Note that $|\rho(1)| \leq 0.5$.

When $\theta_1 = \mathbf{0.5}$ $\Rightarrow \rho(1) = \mathbf{0.4.}$ ($|\theta_1| < 1 \Rightarrow$ invertible)
$\theta_1 = -0.9$ $\Rightarrow \rho(1) = -0.497238.$ ($|\theta_1| < 1 \Rightarrow$ invertible)
$\theta_1 = 2$ $\Rightarrow \rho(1) = \mathbf{0.4.}$ ($|\theta_1| < 1$ non-invertible)

Note: We have two processes, with the same ACF, we select $\theta_1 = \mathbf{0.5.}$

Review: MA Process – MA(1): ACF

Example (continuation):

In general, for an MA(q) process, the k *autocorrelation function* (ACF):

$$\rho(k) = \frac{\sum_{j=k}^{q} \theta_j \theta_{j-k}}{(1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2)} \qquad \text{for } |k| \leq q$$

$$\rho(k) = 0 \qquad \text{for } |k| > q$$

Remark: After lag q, the ACF are 0 (contrast with AR(1) model).

Note: The ACF is usually shown in a plot. When we plot $\rho(k)$ against k, we plot also $\rho(0) = 1$.

Review: MA(1) Process – ACF: Simulations

Example: Below, we compute & plot the ACF for the simulated process.

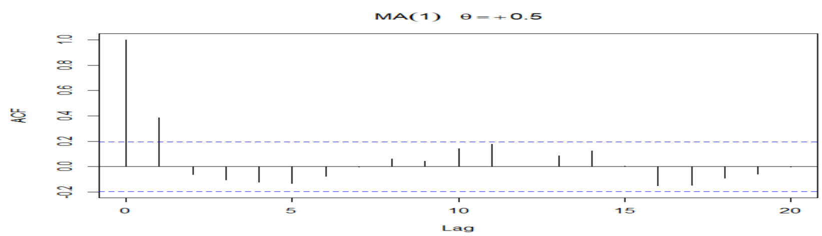
$$y_t = \varepsilon_t + \mathbf{0.5}\, \varepsilon_{t-1}$$

```
sim_ma1_5 <- arima.sim(list(order=c(0,0,1), ma=0.5), n=100)
acf_ma1_5 <- acf(sim_ma1_5, main=(expression(MA(1)~~~theta==+.5)))
> acf_ma1_5
```

Autocorrelations of series 'sim_ma1_5', by lag

0	1	2	3	4	5	6	7	8	9	10	11	12	13
1.000	**0.438**	0.069	0.014	0.103	0.173	0.107	0.015	-0.080	-0.054	0.011	-0.006	0.041	0.000

14	15	16	17	18	19	20	21	22	23
-0.094	-0.147	-0.129	-0.082	-0.150	-0.196	-0.251	-0.235	-0.021	0.110



Review: MA Process – Estimation

• MA processes are more complicated to estimate since we do not observe the errors, ε_t's: Direct estimation is impossible.

• Two indirect ways:
(1) Using method of moments (MM): e matched observed moments and solved for the parameters. For example, for an MA(1):

$$\rho(1) = \theta_1/(1+\theta_1^2)$$

$$r_1 = \frac{\hat{\theta}}{(1+\hat{\theta}^2)} \quad \Rightarrow \quad \hat{\theta} = \frac{1 \pm \sqrt{1-4r_1^2}}{2r_1}$$

• A nonlinear solution and difficult to solve.

(2) Using AR(∞) representation: For MA(1) & $|\theta|<1$, find $a \in (-1; 1)$

$$\varepsilon_t(a) = y_t + a\, y_{t-1} + a^2\, y_{t-2} + a^3\, y_{t-3} + \ldots.$$

and look (numerically) for the least-square estimator

$$\hat{\boldsymbol{\theta}} = \arg\min_\theta \{S(\boldsymbol{y}; \boldsymbol{\theta}) = \textstyle\sum_{i=1}^{T} \varepsilon_i(a)^2\} \qquad (a^i = \theta_1{}^i.)$$

Autoregressive (AR) Process

• We model the conditional expectation of y_t, $E_t[y_t | I_{t-1}]$, as a function of its past history.

• We keep ε_t ~ WN(0, σ^2)

• The most common models are AR models. An AR(1) model involves a single lag, while an AR(p) model involves p lags. Then, the AR(p) process is given by:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t, \qquad \varepsilon_t \sim WN.$$

Using the lag operator we write the AR(p) process: $\phi(L)\, y_t = \varepsilon_t$

with $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$

AR Process: SDE

• We can look at an AR(p) process:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t,$$

as a *stochastic (linear) difference equation* (SDE). With difference equations we try to get a solution –i.e., given some initial conditions/history, we know the value of y_t for any t– and, then, we study its characteristics (stability, long-run value, etc.).

The solution to a DE can be written as a sum of two solutions:

1) Homogeneous equation (the part that only depends on the y_t's):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} \qquad (\text{set } \mu + \varepsilon_t = 0)$$

2) A particular solution to the equation.

• Once we get a solution, we study its stability. We want a stable one.

AR Process – AR(1): Solution

• We get a solution to the simple case, the AR(1) process.

$$y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim WN.$$

Using the backward substitution method:

$$y_t = \mu\,(1 + \phi_1 + \phi_1^2 + ... + \phi_1^{t-1}) + \sum_{j=0}^{t-1} \phi_1^j \varepsilon_{t-j} + \phi_1^t y_0$$

Note: The solution is a function of t, the sequence $\varepsilon_t, \varepsilon_{t-1}, ..., \varepsilon_1$ and initial condition y_0. The effect of y_0 "dies out" if $|\phi_1| < 1$.

• The stability of the solution is crucial. With a stable solution, y_t does not explode. This is good: We need well defined moments.

It turns out the stability of the equation depends on the solution to the homogenous equation. In the AR(1) case (setting μ & ε_t's = 0):

$$y_t = \phi_1^t y_0 \quad \Rightarrow \text{If } |\phi_1| < 1,\ y_t \text{ never explodes, as } t \to \infty.$$

AR Process – AR(1): Solution & Stability

• We can analyze the stability from the point of view of the roots of the lag polynomial. For the AR(1) process

$$\phi(z) = 1 - \phi_1 z = 0 \quad \Rightarrow |z| = \frac{1}{|\phi_1|} > 1$$

That is, the AR(1) process is stable if the root of $\phi(z)$ is greater than one (also said as "**the roots lie outside the unit circle**").

This result generalizes to AR(p) process:

AR Process – AR(2): Solution & Stability

• For the AR(2), $y_t = \phi_1 \, y_{t-1} - \phi_2 \, y_{t-2}$

Lag polynomial: $\phi(L) = 1 - \phi_1 L - \phi_2 \, L^2 = 0.$

We need the roots of $\phi(z)$ to be outside the unit circle.

The characteristic polynomial of the AR(2) can be written as:

$$\phi(z) = 1 - (\lambda_1 + \lambda_2)\, z - \lambda_1 \lambda_2 \, z^2 = (1 - \lambda_1 z)\,(1 - \lambda_2 z) = 0$$

where $\phi_1 = \lambda_1 + \lambda_2$, and $\phi_2 = \lambda_1 \lambda_2$. ($\lambda_1$ & λ_2 are *eigenvalues*.)

If $|\lambda_1|<1$, and $|\lambda_2|<1$, roots lie ***outside the unit root*** $\Rightarrow$ stationary

Then, some implications for ϕ_1 & ϕ_2:

$$|\lambda_1 + \lambda_2| < 2 \quad \Rightarrow |\phi_1| < 2$$
$$|\lambda_1 \lambda_2| < 1 \quad \Rightarrow |\phi_2| < 1$$

AR Process – AR(p): Solution & Stability

• Summary:

We say the process is *globally (asymptotically) stable* if the solution of the associated homogenous equation tends to 0, as $t \to \infty$.

Theorem

A necessary and sufficient condition for global asymptotical stability of a p^{th} order deterministic difference equation with constant coefficients is that **all roots** of the associated lag polynomial equation $\phi(z)$=0 **have *moduli* strictly more than 1**.

(For the case of real roots, ***moduli*** = "**absolute values**.")

AR(1) Process – Stationarity & ACF

- An AR(1) model:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN.$$

Last lecture, under the stationarity condition $|\phi_1| < 1$, we derived:

- **Moments**

$E[y_t] = \mu = 0$ (assuming $\phi_1 \neq 1$)

$Var[y_t] = \gamma(0) = \sigma^2/(1 - \phi_1^2)$ (assuming $|\phi_1| < 1$)

$\gamma(1) = \phi_1 \gamma(0)$

$\gamma(2) = \phi_1^2 \gamma(0)$

$\gamma(3) = \phi_1^3 \gamma(0)$

⋮

$\gamma(k) = \phi_1^k \gamma(0)$

AR(1) Process – Stationarity & ACF

- We derive the autocorrelations:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\phi_1^k \gamma(0)}{\gamma(0)} = \phi_1^k$$

Remark: Assuming $|\phi_1| < 1$, the ACF decays with k.

When we plot $\rho(k)$ against k, we plot also $\rho(0)$ which is 1.

Note:

– when $0 < \phi_1 < 1$ ⇒ All autocorrelations are positive.

– when $-1 < \phi_1 < 0$ ⇒ The sign of $\rho(k)$ shows an alternating sign pattern beginning a negative value.

– when $\phi_1 = 1$ ⇒ AR(1) is non-stationary, $\rho(k) = 1$, for all k. Present & past are always correlated!

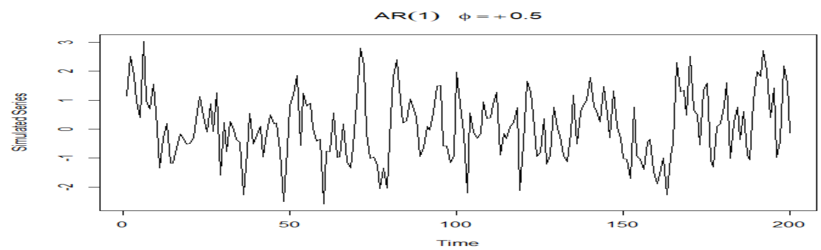
AR(1) Process – Stationarity & ACF: Simulations

Example: We simulate and plot three MA(1) processes, with standard normal ε_t -i.e., σ =1:

$$y_t = 0.5\, y_{t-1} + \varepsilon_t$$
$$y_t = -0.9\, y_{t-1} + \varepsilon_t$$
$$y_t = 2\, y_{t-1} + \varepsilon_t$$

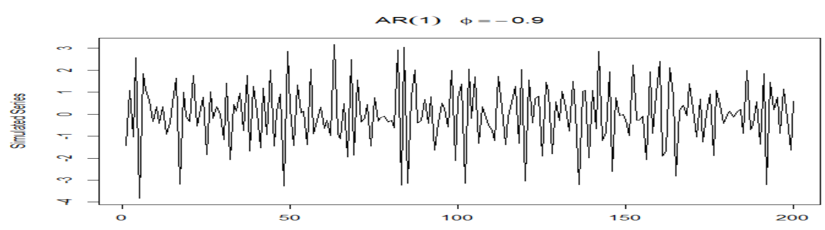
R script to plot y_t = **0.5** y_{t-1} + ε_t with 200 simulations

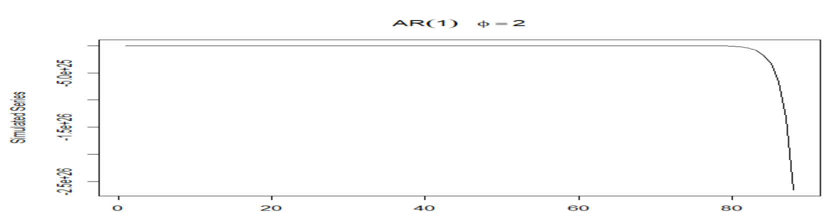
```
> plot(arima.sim(list(order=c(1,0,0), ar=0.5), n=200), ylab="ACF", main=(expression(AR(1)~~~phi==+.5)))
```



AR(1) Process – Stationarity & ACF: Simulations

Example (continuation):





Note: The process $\theta_1 > 0$ is smoother than the ones with $\theta_1 < 0$. The process with $|\theta_1| > 1$, explodes!

AR(1) Process – Stationarity & ACF: Simulations

Example (continuation): Below, we compute and plot the ACF for the the two stable simulated process.

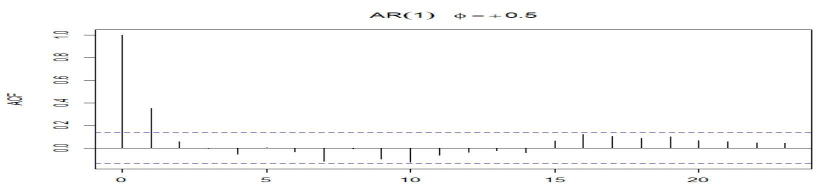
1) $y_t = \mathbf{0.5}\, y_{t-1} + \varepsilon_t$

```
sim_ar1_5 <- arima.sim(list(order=c(1,0,0), ar=0.5), n=200)
acf_ar1_5 <- acf(sim_ar1_5, main=(expression(AR(1)~~~phi==+.5)))
acf_ar1_5
```

Autocorrelations of series 'sim_ma1_5', by lag

0	1	2	3	4	5	6	7	8	9	10	11	12	13
1.000	0.351	0.055	-0.005	-0.054	0.002	-0.036	-0.119	-0.008	-0.099	-0.125	-0.066	-0.036	-0.023

14	15	16	17	18	19	20	21	22	23
-0.042	0.062	0.119	0.102	0.087	0.099	0.065	0.056	0.047	0.044



AR(1) Process – Stationarity & ACF: Simulations

Example (continuation):

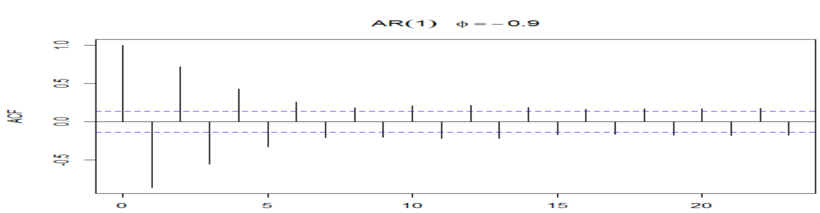
2) $y_t = \mathbf{-\ 0.9}\, y_{t-1} + \varepsilon_t$

```
sim_ar1_9 <- arima.sim(list(order=c(1,0,0), ar=-0.9), n=200) acf_ar1_9 <-
acf(sim_ar1_9, main=(expression(AR(1)~~~phi==-.9)))
> acf_ar1_9
```

Autocorrelations of series 'sim_ma1_9', by lag

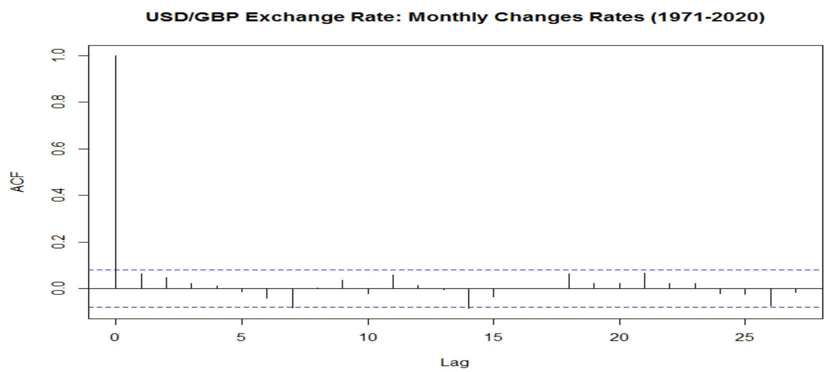
0	1	2	3	4	5	6	7	8	9	10	11	12	13
1.000	-0.584	0.093	0.061	-0.132	0.147	-0.181	0.122	-0.013	-0.023	0.014	-0.012	0.092	-0.199

14	15	16	17	18	19	20	21	22	23
0.193	-0.155	0.143	-0.107	0.014	0.174	-0.244	0.196	-0.154	0.105



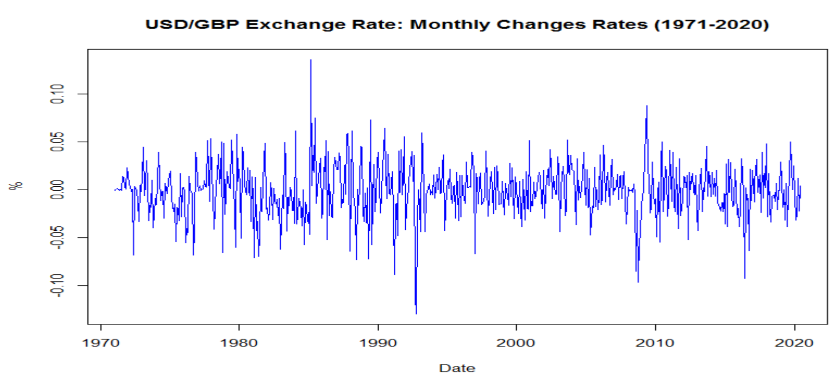
AR(1) Process – Stationarity & ACF: Examples

Example: A process with $|\phi_1| < 1$ (actually, 0.065) is the **monthly changes in the USD/GBP exchange rate**. Below we plot its corresponding ACF:



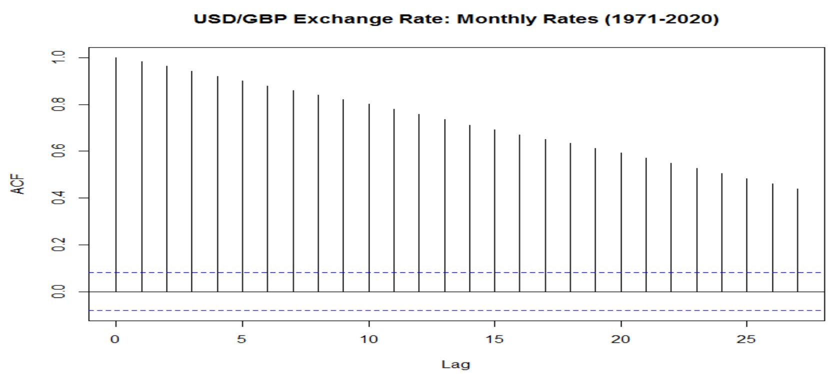
AR(1) Process – Stationarity & ACF

Example: Below we plot the monthly **changes in the USD/GBP exchange rate**. Stationary series do not look smooth:



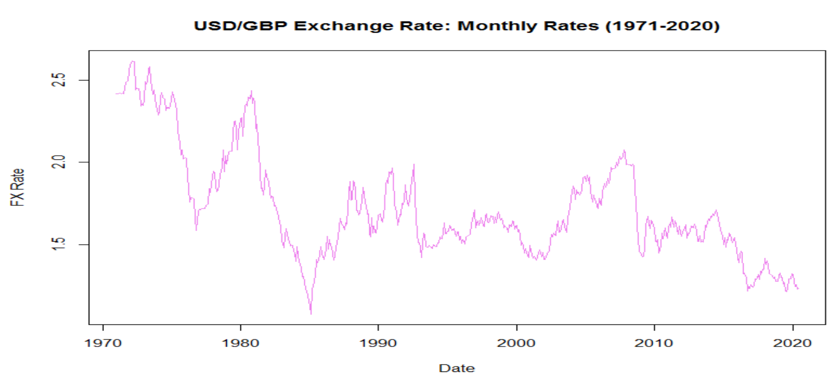
AR(1) Process – Stationarity & ACF

Example: A process with $\phi_1 \approx 1$ (actually, 0.99) is the **nominal USD/GBP exchange rate**. Below, we plot the ACF, it is not 1 all the time, but its decay is very slow (after 30 months, it is still .40 correlated!):



AR(1) Process – Stationarity & ACF

Example: Below we plot the **nominal USD/GBP exchange rate**. Stationary series look smooth, smooth enough that you can clearly spot trends:



AR(2) Process – Stationarity & ACF

- An AR(2) model:

$$y_t = \mu + \phi_1\, y_{t-1} + \phi_2\, y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim WN.$$

- **Moments:** (μ =0)

$$E[y_t] = \frac{\mu}{(1-\phi_1-\phi_2)} = 0 \qquad \text{(assuming } \phi_1 + \phi_2 \neq 1\text{)}$$

$$Var[y_t] = \frac{\sigma^2}{(1-\phi_1^2-\phi_2^2)} \qquad \text{(assuming } \phi_1^2 + \phi_2^2 < 1\text{)}$$

- **Autocovariance function**

$$\gamma(k) = Cov[y_t, y_{t-k}] = E[(\phi_1\, y_{t-1} + \phi_2\, y_{t-2} + \varepsilon_t)\, y_{t-k}]$$
$$= \phi_1\, E[y_{t-1}\, y_{t-k}] + \phi_2\, E[y_{t-2}\, y_{t-k}] + E[\varepsilon_t\, y_{t-k}]$$
$$= \phi_1\, \gamma(k-1) + \phi_2\, \gamma(k-2) + E[\varepsilon_t\, y_{t-k}]$$

We have a recursive formula.

AR(2) Process – Stationarity & ACF

- Recursive formula: $\gamma(k) = \phi_1\, \gamma(k-1) + \phi_2\, \gamma(k-2) + E[\varepsilon_t\, y_{t-k}]$

(k=0) $\gamma(0) = \phi_1\, \gamma(-1) + \phi_2\, \gamma(-2) + E[\varepsilon_t\, y_t]$
$\quad = \phi_1\, \gamma(1) + \phi_2\, \gamma(2) + \sigma^2$

(k=1) $\gamma(1) = \phi_1\, \gamma(0) + \phi_2\, \gamma(1) + E[\varepsilon_t\, y_{t-1}]$
$\quad = \phi_1\, \gamma(0) + \phi_2\, \gamma(1) + 0$
$\Rightarrow \gamma(1) = [\phi_1/(1-\phi_2)]\ \gamma(0)$

(k=2) $\gamma(2) = \phi_1\, \gamma(1) + \phi_2\, \gamma(0) + E[\varepsilon_t\, y_{t-2}]$
$\quad = \phi_1\, \gamma(1) + \phi_2\, \gamma(0) + 0$
$\Rightarrow \gamma(2) = [\phi_1^2/(1-\phi_2) + \phi_2]\ \gamma(0)$

Replacing $\gamma(1)$ and $\gamma(2)$ back to $\gamma(0)$:

$$\gamma(0) = [\phi_1^2/(1 - \phi_2)]\ \gamma(0) + [\phi_2\, \phi_1^2/(1 - \phi_2) + \phi_2^2]\ \gamma(0) + \sigma^2$$

$$= \frac{\sigma^2(1-\phi_2)}{(1-\phi_2) - \phi_1^2\,(1+\phi_2) + \phi_2^2\,(1-\phi_2)} \qquad \Rightarrow |\phi_2|<1$$

AR(2) Process – Stationarity & ACF

- Dividing the recursive formula for $\gamma(k)$ by $\gamma(0)$, we get the ACF:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \phi_1\,\rho(k-1) + \phi_2\,\rho(k-2) \;+ \frac{E[\varepsilon_t\, y_{t-k}]}{\gamma(0)}$$

(k=0) $\rho(0) = 1$

(k=1) $\rho(1) = \phi_1/(1-\phi_2)$

(k=2) $\rho(2) = \phi_1\,\rho(1) + \phi_2\,\rho(0) = \phi_1^2/(1-\phi_2) + \phi_2$

(k=3) $\rho(3) = \phi_1\,\rho(2) + \phi_2\,\rho(1) =$
$= \phi_1^3/(1-\phi_2) + \phi_1\,\phi_2 + \phi_2\,\phi_1/(1-\phi_2)$

<u>Remark</u>: Again, we see exponential decay in the ACF.

From the work above, for stationarity, we need: $\phi_1 + \phi_2 \neq 1$.
$\phi_1^2 + \phi_2^2 < 1$.
$|\phi_2| < 1$.

AR Process – Stationarity and Ergodicity

Theorem: The linear AR(p) process is strictly stationary and ergodic if and only if the roots of ϕ(L) are $|z_j|$>1 for all j, where $|z_j|$ is the modulus of the complex number r_j.

<u>Note</u>: If one of the z_j's equals 1, ϕ(L) (& y_t) has a **unit root** –i.e., ϕ(1)=0. This is a special case of *non-stationarity*.

- Recall $\phi(L)^{-1}$ produces an infinite sum on the ε_{t-j}'s. If this sum does not explode, we say the process is **stable**.

- If the process is stable, we can calculate $\delta y_t/\delta\varepsilon_{t-j}$.

$\frac{\delta y_t}{\delta \varepsilon_{t-j}}$ = How much y_t is affected today by an innovation $t-j$ periods ago, ε_{t-j}.

When expressed as a function of j, we call this *dynamic multiplier*.

AR Process – Dynamic Multiplier

• The *dynamic multiplier* measurers the effect of an innovation, ε_t, (economist like to call the ε_t's, "*shocks*") on subsequent values of y_t: That is, the first derivative on the "**Wold representation**" –i.e., a stationary process represented as an MA process:

$$\frac{\delta y_{t+j}}{\delta \varepsilon_t} = \frac{\delta y_t}{\delta \varepsilon_0} = \psi_j.$$

where ψ_j's are the coefficient of the (inverted) AR representation.

For an AR(1) process:

$$\frac{\delta y_{t+j}}{\delta \varepsilon_t} = \frac{\delta y_t}{\delta \varepsilon_0} = \phi^j$$

• That is, the dynamic multiplier for any linear stochastic difference equation (SDE) depends only on the length of time j, not on time t.

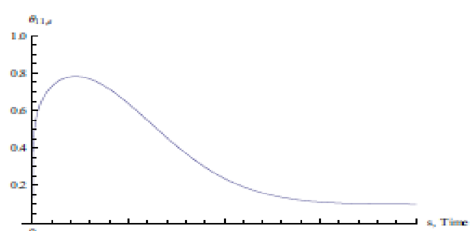
AR Process – Impulse Response Function

• The **impulse-response function** (**IRF**) is an accumulation of the sequence of dynamic multipliers, as a function of time from the one time change in the innovation, ε_t.

• Usually, IRFs are represented with a graph, that measures the effect of the innovation, ε_t, on y_t over time:

$$\frac{\delta y_{t+j}}{\delta \varepsilon_t} + \frac{\delta y_{t+j+1}}{\delta \varepsilon_t} + \frac{\delta y_{t+j+2}}{\delta \varepsilon_t} + \ldots = \psi_j + \psi_{j+1} + \psi_{j+2} + \ldots$$

• Once we estimate the AR, MA or ARMA coefficients, we draw an IRF.



AR Process – IRF: AR(1)

Example: AR(1) process:

$$y_t = \mu + \phi_1\, y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim WN.$$

The AR(1) is stable if $|\phi_1|<1$ ⇒ stationarity condition.

We invert the AR(1) to get an MA(∞): $1/(1-\phi_1) = \sum_{j=0}^{\infty} \phi_1^j$

Then,

$$y_t = \mu* + \phi_1^1\, \varepsilon_{t-1} + \phi_2^2\, \varepsilon_{t-2} + \phi_2^3\, \varepsilon_{t-3} + \phi_2^4\, \varepsilon_{t-4} + \cdots + \varepsilon_t.$$

Under the stationarity condition, we calculate the dynamic multiplier:

$$\delta y_{t+1}/\delta\varepsilon_{t-j} = \phi_1^j$$

Accumulated over time, after J periods, the effect of shock ε_t at t+J is:

$$\text{IRF(at } t+J) = \sum_{j=0}^{J-1} \phi_1^j$$

AR Process – IRF: AR(1)

Example (continuation): Suppose ϕ_1= **0.40**. Then,

$$\delta y_t/\delta\varepsilon_{t-1} = \phi_1 = \mathbf{0.40}$$
$$\delta y_t/\delta\varepsilon_{t-2} = \phi_1^2 = \mathbf{0.40}^2$$
$$\vdots$$
$$\delta y_t/\delta\varepsilon_{t-J} = \phi_1^J = \mathbf{0.40}^J$$

After $J = 5,$ periods, the accumulated effect of a shock today is:

$$\text{IRF(at } t+5) = \mathbf{0.40} + \mathbf{0.40}^2 + \mathbf{0.40}^3 + \mathbf{0.40}^4 + \mathbf{0.40}^5 = 0.65984$$

AR Process – Estimation and Properties

• We go back to the general AR(p). Define

$$\boldsymbol{x}_t = (1\ y_{t-1}\ y_{t-2} \ldots .\ y_{t-p})$$
$$\boldsymbol{\beta} = (\mu\ \phi_1\ \phi_2\ \ldots .\ \phi_p)$$

Then the model can be written as

$$y_t = \boldsymbol{x}_t{}'\boldsymbol{\beta} + \varepsilon_t$$

• The OLS estimator is $\mathbf{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$

• Properties:
  – Using the Ergodic Theorem, OLS estimator is consistent.
  – Using the MDS CLT, OLS estimator is asymptotically normal.
    ⇒ asymptotic inference is the same.

• The asymptotic covariance matrix is estimated just as in the cross-section case: The sandwich estimator.

ARMA Process

• A combination of AR(p) and MA(q) processes produces an ARMA(p, q) process:

$$y_t = \mu + \phi_1\, y_{t-1} + \phi_2\, y_{t-2} + \ldots + \phi_p\, y_{t-p} + \varepsilon_t - \theta_1\, \varepsilon_{t-1} - \theta_2\, \varepsilon_{t-2} - \ldots - \theta_q\, \varepsilon_{t-q}$$
$$= \mu + \sum_{i=1}^{p} \phi_i\, y_{t-i} - \sum_{i=1}^{q} \theta_i L^i \varepsilon_t + \varepsilon_t$$

$$\Rightarrow \phi(L) y_t = \mu + \theta(L)\varepsilon_t$$

• Usually, we insist that $\phi(L) \neq 0$, $\theta(L) \neq 0$ & that the polynomials $\phi(L)$, $\theta(L)$ have no *common factors*. This implies it is not a lower order ARMA model.

ARMA(1,1) – Stationarity & ACF

- For an ARMA(1,1) we have:.

$$y_t = \mu + \phi_1\, y_{t-1} + \theta_1\, \varepsilon_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim WN.$$

- **Moments:** ($\mu = 0$)

$E[y_t] = \mu \,/\, (1 - \phi_1) = 0$ (assuming $\phi_1 \neq 1$)

$Var[y_t] = \sigma^2\, (1 + \theta_1^2)\, /\, (1 - \phi_1^2)$ (assuming $|\phi_1| < 1$)

- **Autocovariance function** ($\mu = 0$)

$$\begin{aligned}\gamma(k) &= Cov[y_t,\ y_{t-k}] \\ &= E[\{\phi_1\, y_{t-1} + \theta_1\, \varepsilon_{t-1} + \varepsilon_t\}\, y_{t-k}] \\ &= \phi_1\, E[y_{t-1}\, y_{t-k}] + \theta_1\, E[\varepsilon_{t-1}\, y_{t-k}] + E[\varepsilon_t\, y_{t-k}] \\ &= \phi_1\, \gamma(k-1) + \theta_1\, E[\varepsilon_{t-1}\, y_{t-k}] + E[\varepsilon_t\, y_{t-k}]\end{aligned}$$

- Again, we have a recursive formula.

$$\gamma(k) = \phi_1\, \gamma(k-1) + \theta_1\, E[\varepsilon_{t-1}\, y_{t-k}] + E[\varepsilon_t\, y_{t-k}]$$

ARMA(1,1) – Stationarity & ACF

- We have a recursive formula:

$$\gamma(k) = \phi_1 \gamma(k-1) + E[\varepsilon_t\, y_{t-k}] + \theta_1\, E[\varepsilon_{t-1}\, y_{t-k}]$$

For $k = 0$,

$$\begin{aligned}\gamma(0) &= \phi_1\, \gamma(-1) + \underbrace{E[\varepsilon_t\, y_t]}_{\sigma^2} + \theta_1\, E\Big[\varepsilon_{t-1}\ \underbrace{y_t}_{\phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}}\Big] \\ &= \phi_1\, \gamma(1) + \sigma^2 + \theta_1\, E\Big[\varepsilon_{t-1}(\phi_1\, \underbrace{y_{t-1}}_{\phi_1 y_{t-2} + \varepsilon_{t-1} + \theta_1 \varepsilon_{t-2}} + \varepsilon_t + \theta_1 \varepsilon_{t-1})\Big] \\ &= \phi_1\, \gamma(1) + \sigma^2 + \theta_1\, (\phi_1\, \sigma^2 + \theta_1 \sigma^2)\end{aligned}$$

For $k = 1$,

$$\begin{aligned}\gamma(1) &= \phi_1 \gamma(0) + E[\varepsilon_t\, y_{t-1}] + \theta_1 E[\varepsilon_{t-1}\, y_{t-1}] \\ &= \phi_1 \gamma(0) + \theta_1\, E[\varepsilon_{t-1}\, \{\phi_1 y_{t-2} + \theta_1 \varepsilon_{t-2} + \varepsilon_{t-1}\}] \\ &= \phi_1 \gamma(0) + \theta_1\, \gamma(1)\end{aligned}$$

ARMA(1,1) – Stationarity & ACF

- $\gamma(k) = \phi_1\gamma(k-1) + E[\varepsilon_t\, y_{t-k}] + \theta_1\, E[\varepsilon_{t-1}\, y_{t-k}]$

For k = 2,

$$\begin{aligned}\gamma(2) &= \phi_1\gamma(1) + E[\varepsilon_t\, y_{t-2}] + \theta_1\, E[\varepsilon_{t-1}\, y_{t-2}] \\ &= \phi_1\gamma(1) + \theta_1\, E[\varepsilon_{t-1}\, \{\phi_1\, y_{t-3} + \theta_1\, \varepsilon_{t-3} + \varepsilon_{t-2}\}] \\ &= \phi_1\gamma(1)\end{aligned}$$

For k,

$$\begin{aligned}\gamma(k) &= \phi_1\,\gamma(k-1) \\ &= \phi_1^{k-1}\,\gamma(1), \quad k > 1\end{aligned}$$

$\Rightarrow$ If $|\phi_1|$<1, exponential decay.

- Two equations for $\gamma(0)$ and $\gamma(1)$:

$$\gamma(0) = \phi_1\,\gamma(1) + \sigma^2 + \theta_1\,(\phi_1\,\sigma^2 + \theta_1\,\sigma^2)$$

$$\gamma(1) = \phi_1\,\gamma(0) + \theta_1\,\gamma(1)$$

ARMA(1,1) – Stationarity & ACF

- Two equations for $\gamma(0)$ and $\gamma(1)$:

$$\gamma(0) = \phi_1\,\gamma(1) + \sigma^2 + \theta_1\,(\phi_1\,\sigma^2 + \theta_1\,\sigma^2)$$

$$\gamma(1) = \phi_1\,\gamma(0) + \theta_1\,\gamma(1)$$

Solving for $\gamma(0)$ & $\gamma(1)$:

$$\gamma(0) = \sigma^2\,\frac{1+\theta_1{}^2 + 2\,\phi_1\theta_1}{1-\phi_1{}^2}$$

$$\gamma(1) = \sigma^2\,\frac{(1+\phi_1\theta_1)*(\phi_1+\theta_1)}{1-\phi_1{}^2}$$

$\vdots$

$\gamma(k) = \phi_1{}^{k-1}\,\gamma(1), \quad k > 1$ $\Rightarrow$ If $|\phi_1|$<1, exponential decay.

Note: If stationary, ARMA(1,1) & AR(1) show exponential decay. Difficult to distinguish one from the other through autocovariances.

ARMA Process – Common Factors

Example: Common factors.
Suppose we have the following ARMA(2, 3) model

$$\phi(L)y_t = \theta(L)\varepsilon_t$$

with

$$\phi(L) = 1 - .6L + .3L^2$$
$$\theta(L) = 1 - 1.4L + .9L^2 - .3L^3 = (1 - .6L + .3L^2)(1 - L)$$

This model simplifies to: $y_t = (1 - L)\varepsilon_t$ $\Rightarrow$ an MA(1) process.

- We just simplify the common factors and keep the simpler representation.

ARMA Process – Representation

- AR Representation: $\Pi(L)(y_t - \mu) = \varepsilon_t \Rightarrow \Pi(L) = \frac{\phi_p(L)}{\theta_q(L)}$

- Pure MA Representation: $(y_t - \mu) = \Psi(L)\varepsilon_t \Rightarrow \Psi(L) = \frac{\theta_q(L)}{\phi_p(L)}$

- Special ARMA(p, q) cases:
  – p = 0: MA(q)
  – q = 0: AR(p).

ARMA: Stationarity, Causality and Invertibility

Theorem: If ϕ(L) and θ(L) have no common factors, a (unique) *stationary* solution to $\phi(L)y_t = \theta(L)\varepsilon_t$ exists if and only if

$$|z| \leq 1 \Rightarrow \phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p \neq 0.$$

This ARMA(p, q) model is causal if and only if

$$|z| \leq 1 \Rightarrow \phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p \neq 0.$$

This ARMA(p, q) model is invertible if and only if

$$|z| \leq 1 \Rightarrow \theta(z) = 1 + \theta_1 z - \theta_2 z^2 + \ldots + \theta_p z^p \neq 0.$$

Note: Real data cannot be *exactly* modeled using a finite number of parameters. We choose p, q to create a good approximated model.

ARMA Process

• We defined the ARMA(p, q) model:

$$\phi(L)(y_t - \mu) = \theta(L)\varepsilon_t$$

The mean does not affect the order of the ARMA. Then, if $\mu \neq 0$, we demean the data: $x_t = y_t - \mu$.

Then, $\phi(L)\ x_t = \theta(L)\ \varepsilon_t \quad \Rightarrow x_t$ is a *demeaned* ARMA process.

• For the rest of the lecture, we will study:
- Identification of p, q.
- Estimation of ARMA(p, q)

Autocovariance Function (Again)

- We define the autocovariance function: $\gamma(t-j) = E[y_t\, y_{t-j}]$

- For an AR(p) process, WLOG with μ=0 (or demeaned y_t), we get:

$$\gamma(t-j) = E[(\phi_1\, y_{t-1}y_{t-j} + \phi_2 y_{t-2}y_{t-j} + \ldots + \phi_p y_{t-p}y_{t-j} + \varepsilon_t y_{t-j})]$$
$$= \phi_1\, \gamma(j-1) + \phi_2 \gamma(j-2) + \ldots + \phi_p \gamma(j-p)$$

- The autocovariances, $\gamma(t-j)$, determine a system of equations:

$$\gamma(0) = E[y_t, y_t] = \phi_1\, \gamma(1) + \phi_2\, \gamma(2) + \phi_3\, \gamma(3) + \ldots + \phi_p\, \gamma(p) + \sigma^2$$
$$\gamma(1) = E[y_t, y_{t-1}] = \phi_1\, \gamma(0) + \phi_2\, \gamma(1) + \phi_3\, \gamma(2) + \ldots + \phi_p\, \gamma(p-1)$$
$$\gamma(2) = E[y_t, y_{t-2}] = \phi_1\, \gamma(1) + \phi_2\, \gamma(0) + \phi_3\, \gamma(1) + \ldots + \phi_p\, \gamma(p-2)$$
$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

Autocovariance Function (Again)

- The pxp system of equations:

$$\gamma(1) = E[y_t, y_{t-1}] = \phi_1\gamma(0) + \phi_2\gamma(1) + \phi_3\gamma(2) + \ldots + \phi_p\gamma(p-1)$$
$$\gamma(2) = E[y_t, y_{t-2}] = \phi_1\gamma(1) + \phi_2\gamma(0) + \phi_3\gamma(1) + \ldots + \phi_p\gamma(p-2)$$
$$\gamma(3) = E[y_t, y_{t-3}] = \phi_1\gamma(2) + \phi_2\gamma(1) + \phi_3\gamma(0) + \ldots + \phi_p\gamma(p-3)$$
$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

Using linear algebra, we write the system as: $\boldsymbol{\gamma} = \boldsymbol{\Gamma}\, \boldsymbol{\phi}$

where

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(p-2) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \cdots & \gamma(0) \end{bmatrix} \quad \text{a } p\text{x}p \text{ matrix}$$

$\boldsymbol{\phi}$ is the px1 vector of AR(p) coefficients
$\boldsymbol{\gamma}$ is the px1 vector of $\gamma(k)$ autocovariances.

Autocorrelation Function (ACF)

- Now, we define the autocorrelation function (**ACF**):

$$\rho(k) = \frac{\gamma(k))}{\gamma(0)} = \frac{\text{covariance at lag } k}{\text{variance}}$$

The ACF lies between -1 and +1, with $\rho(0) = 1$.

- Dividing the autocovariance system by γ(0), we get:

$$\begin{bmatrix} \rho(0) & \rho(1) & \cdots & \rho(p-1) \\ \rho(1) & \rho(0) & \cdots & \rho(p-2) \\ \vdots & \vdots & \cdots & \vdots \\ \rho(p-1) & \rho(p-2) & \cdots & \rho(0) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(p) \end{bmatrix}$$

Or using linear algebra: $\mathbf{P}\,\boldsymbol{\phi} = \boldsymbol{\rho}$

- These are **Yule-Walker** equations, which can be solved numerically.

ACF – Estimation & Correlogram

- **Estimation**:

Easy: Use sample moments to estimate γ(*k)* and plug in formula:

$$r_k = \hat{\rho}_k = \frac{\sum (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum (Y_t - \bar{Y})^2}$$

Then, we plug the $\hat{\rho}_k$ in the Yule-Walker equations and solve for $\boldsymbol{\phi}$:

$$\widehat{\mathbf{P}}\,\boldsymbol{\phi} = \widehat{\boldsymbol{\rho}}$$

- The sample *correlogram* is the plot of the ACF against *k*. As the ACF lies between -1 and +1, the correlogram also lies between these values.

:

ACF – Distribution

• **Distribution**:
For a linear, stationary process, with large T, the distribution of the sample ACF, $r_k = \hat{\rho}_k$ is approximately normal with:

$$\mathbf{r} \xrightarrow{d} \text{N}(\boldsymbol{\rho}, \mathbf{V}/T), \qquad \mathbf{V} \text{ is the covariance matrix.}$$

Under H_0: $\rho_k = 0$ for all $k > 1$.

$$\mathbf{r} \xrightarrow{d} \text{N}(\mathbf{0}, \mathbf{I}/T) \quad \Rightarrow \text{Var}[r(k)] = 1/T.$$

• Under H_0, the SE = $1/\sqrt{T}$ $\quad$ **⇒ 95% C.I.: 0 $\pm$ 1.96 * $1/\sqrt{T}$**

Then, for a white noise sequence, approximately 95% of the sample ACFs should be within the above C.I. limits.

Note: The SE = $1/\sqrt{T}$ are sometimes referred as *Bartlett's SE*.

ACF – AR(1)

Example: Sample ACF for an AR(1) process:
Under stationarity:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \phi_1^k \qquad k = 0, 1, 2, \ldots$$

If $|\phi_1| < 1$, the ACF will show exponential decay.

• Suppose $\phi_1 = 0.4$. Then,

$\rho(0) = \mathbf{1}$
$\rho(1) = \mathbf{0.4}$
$\rho(2) = 0.4^2 = \mathbf{0.16}$
$\rho(3) = 0.4^3 = \mathbf{0.064}$
$\rho(4) = 0.4^4 = \mathbf{0.0256}$
⋮
$\rho(k) = \mathbf{0.4^k}$

ACF – AR(1)

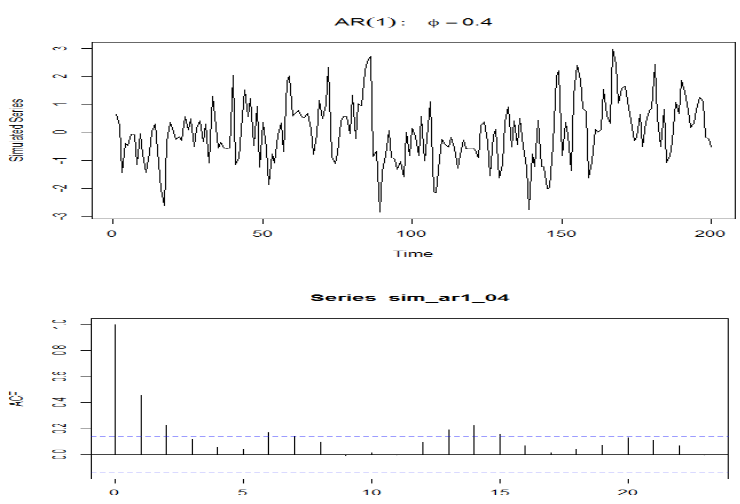
Example (continuation): $\rho(k) = 0.4^k$

We simulate an AR(1) series with with ϕ_1 = 0.4, using the R function *arima.sim*.

```
sim_ar1_04 <- arima.sim(list(order=c(1,0,0), ar=0.4), n=200)          #simulate AR(1) series
plot(sim_ar1_04, ylab="Simulated Series", main=(expression(AR(1):~~~phi==0.4)))
acf(sim_ar1_04)                                                        #plot ACF for sim series
```

ACF – AR(1)

Example (continuation): Plot of simulated series and ACF



ACF – MA(1)

Example (continuation): Sample ACF for an MA(1) process.

$\rho(0) = 1$

$\rho(1) = \theta_1/(1+\theta_1^2)$ for k = 1, -1

$\rho(k) = 0$ for $|k| > 1$.

After k = 1 –i.e., one lag– the ACF dies out.

Suppose θ_1 = **0.5**. Then,

$\rho(0)$ = **1**

$\rho(1)$ = **0.4**

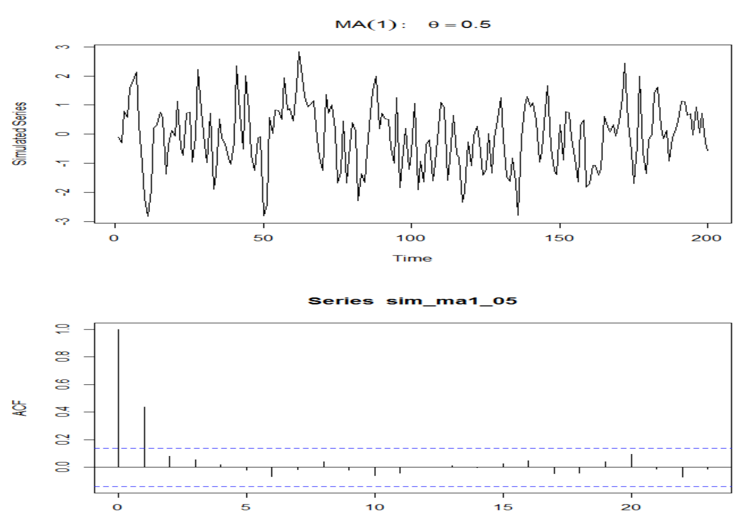
$\rho(k)$ = **0** for $|k| > 1$.

We simulate an MA(1) series with ϕ_1=0.4

```
sim_ma1_05 <- arima.sim(list(order=c(0,0,1), ma=0.5), n=200)        #simulate MA(1) series
plot(sim_ma1_05, ylab="Simulated Series", main=(expression(MA(1):~~~theta==0.5)))
acf(sim_ma1_05)                                                    #plot ACF for sim series
```

ACF – MA(1)

Example (continuation): Plot of simulated series and ACF



ACF – MA(q)

Example: Sample ACF for an MA(q) process:

$$y_t = \mu + \varepsilon_t + \theta_1\ \varepsilon_{t-1} + \theta_2\ \varepsilon_{t-2} + \ldots + \theta_q\ \varepsilon_{t-q}$$

$$\rho(k) = \frac{\sum_{j=k}^{q} \theta_j \theta_{j-k}}{(1+\theta_1{}^2 + \theta_2{}^2 + \ldots + \theta_q{}^2)} \qquad k \le q$$

$$= 0 \qquad \text{otherwise.}$$

For different k's:

$\rho(0) = 1$

$\rho(1) = \frac{\theta_1 + \theta_2\theta_1 + \theta_3\theta_2}{(1+\theta_1{}^2 + \theta_2{}^2 + \theta_3{}^2)}$

$\rho(2) = \frac{\theta_2 + \theta_3\theta_1}{(1+\theta_1{}^2 + \theta_2{}^2 + \theta_3{}^2)}$

$\rho(3) = \frac{\theta_3}{(1+\theta_1{}^2 + \theta_2{}^2 + \theta_3{}^2)}$

$\rho(k) = 0$ for $|k| > 3$.

ACF – MA(q=3)

Example (continuation):

$$y_t = \mu + \varepsilon_t + \theta_1\ \varepsilon_{t-1} + \theta_2\ \varepsilon_{t-2} + \theta_3\ \varepsilon_{t-3}$$

Suppose $\theta_1 = 0.5$; $\theta_2 = 0.4$; $\theta_3 = 0.2$. Then,

$\rho(0) = 1$

$\rho(1) = \frac{\theta_1 + \theta_2\theta_1 + \theta_3\theta_2}{(1+\theta_1{}^2 + \theta_2{}^2 + \theta_3{}^2)} = \frac{0.5+0.4*0.5+0.1*0.4}{1+0.5^2+0.4^2+0.1^2} = \mathbf{0.5211}$

$\rho(2) = \frac{\theta_2 + \theta_3\theta_1}{(1+\theta_1{}^2 + \theta_2{}^2 + \theta_3{}^2)} = \frac{0.4+0.1*0.5}{1+0.5^2+0.4^2+0.1^2} = \mathbf{0.3169}$

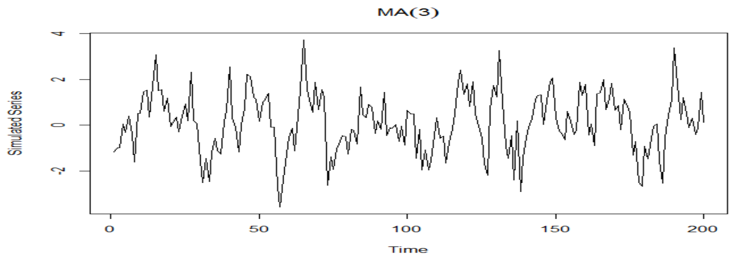
$\rho(3) = \frac{\theta_3}{(1+\theta_1{}^2 + \theta_2{}^2 + \theta_3{}^2)} = \frac{0.1}{1+0.5^2+0.4^2+0.1^2} = \mathbf{0.0704}$

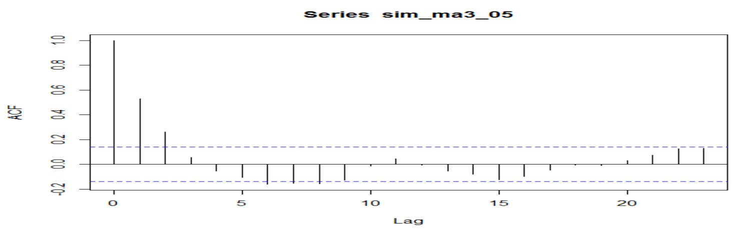
$\rho(k) = \mathbf{0}$ for $|k| > 3$.

ACF – MA(q=3)

Example (continuation): Plot of simulated series and ACF

```
> sim_ma3_05 <- arima.sim(list(order=c(0,0,3), ma=c(0.5, 0.4, 0.1)), n=200)      # sim MA(3)
```





ACF – ARMA(1,1)

Example: Sample ACF for an ARMA(1,1) process:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

• From the autocovariances, we get

$$\gamma(0) = \sigma^2 \frac{1 + \theta_1^{\ 2} + 2\phi_1 \theta_1}{1 - \phi_1^{\ 2}}$$

$$\gamma(1) = \sigma^2 \frac{(1 + \phi_1 \theta_1) * (\phi_1 + \theta_1)}{1 - \phi_1^{\ 2}}$$

$$\gamma(k) = \phi_1 \gamma(k-1) = \phi_1^{\ k-1} \sigma^2 \frac{(1 + \phi_1 \theta_1) * (\phi_1 + \theta_1)}{1 - \phi_1^{\ 2}}$$

• Then,

$$\rho(k) = \phi_1^{\ k-1} \frac{(1 + \phi_1 \theta_1) * (\phi_1 + \theta_1)}{1 + \theta_1^{\ 2} + 2\phi_1 \theta_1}$$

⇒ If $|\phi_1|<1$, exponential decay. Similar pattern to AR(1).

ACF – ARMA(1,1)

Example (continuation): Sample ACF for an ARMA(1,1) process:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

The ACF for an ARMA(1,1):

$$\rho(k) = \phi_1^{k-1} \frac{(1 + \phi_1 \theta_1) * (\phi_1 + \theta_1)}{1 + \theta_1^2 + 2\phi_1\theta_1}$$

- Suppose $\phi_1 = \mathbf{0.4}$, $\theta_1 = \mathbf{0.5}$. Then,

$$\rho(0) = \mathbf{1}$$

$$\rho(1) = \frac{(1 + 0.4 * 0.5) * (0.4 + 0.5)}{1 + 0.5^2 + 2*0.4*0.5} = \mathbf{0.6545}$$

$$\rho(2) = 0.4 * \frac{(1 + 0.4 * 0.5) * (0.4 + 0.5)}{1 + 0.5^2 + 2*0.4*0.5} = \mathbf{0.2618}$$

$$\rho(3) = 0.4^2 * \frac{(1 + 0.4 * 0.5) * (0.4 + 0.5)}{1 + 0.5^2 + 2*0.4*0.5} = \mathbf{0.0233}$$

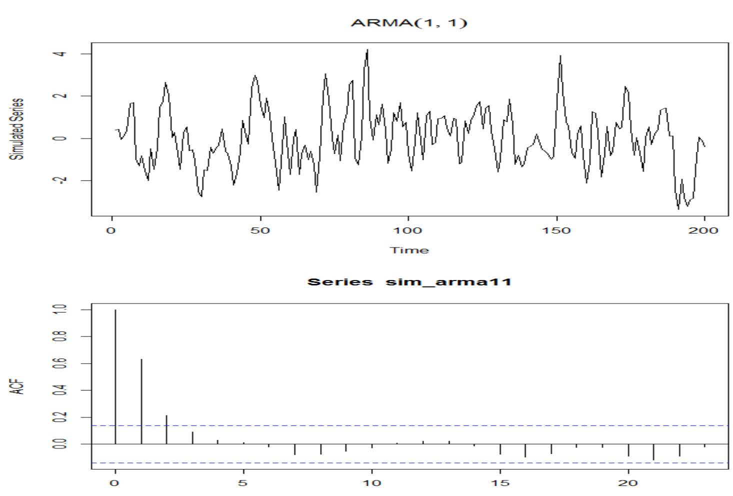
⋮

$$\rho(k) = 0.4^{k-1} * \frac{(1 + 0.4 * 0.5) * (0.4 + 0.5)}{1 + 0.5^2 + 2*0.4*0.5}$$

ACF – ARMA(1,1)

Example (continuation): Plot of simulated series and ACF

```
> sim_arma11 <- arima.sim(list(order=c(1,0,1), ar=0.4, ma=0.5), n=200)    #sim ARMA(1,1)
```



ACF – Example: U.S. Stock Returns

Example: **US Monthly Returns** (1871 – 2020, T=1,795)

```
Sh_da <- read.csv("C://Financial Econometrics/Shiller_2020data.csv", head=TRUE, sep=",")
x_P <- Sh_da$P
x_D <- Sh_da$D
T <- length(x_P)
lr_p <- log(x_P[-1]/x_P[-T])
lr_d <- log(x_D[-1]/x_D[-T])
acf_p <- acf(lr_p)                    # acf: R function that estimates the ACF
> acf_p
```

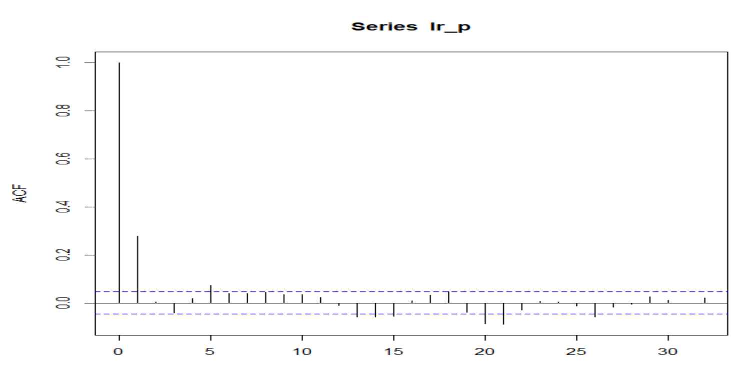
Autocorrelations of series 'lr_p', by lag

0	1	2	3	4	5	6	7	8	9	10	11
1.000	0.279	0.004	-0.043	0.017	0.074	0.039	0.039	0.044	0.035	0.034	0.022
12	13	14	15	16	17	18	19	20	21	22	23
-0.010	-0.059	-0.058	-0.056	0.009	0.033	0.047	-0.040	-0.087	-0.090	-0.029	0.005
24	25	26	27	28	29	30	31	32			
0.003	-0.013	-0.058	-0.018	-0.005	0.026	0.011	0.000	0.020			

SE(r_k) = 1/sqrt(T) = 1/sqrt(1,795) = .0236. ⇒ 95% C.I.: ± 2* 0.0236

ACF – Example: U.S. Stock Returns

Example (continuation): Correlogram for **US Monthly Returns** (1871 – 2020)



Note: With the exception of first correlation, correlations are small. However, many are significant, not strange result when T is large.

ACF – Example: U.S. Stock Dividends

Example: **US Monthly Changes in Dividends** (1871 – 2020, T=1,795)

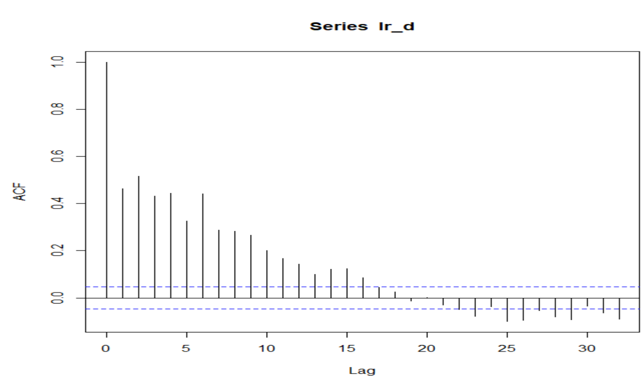
```
acf_d <- acf(lr_d)
> acf_d
Autocorrelations of series ‘lr_d’, by lag
```

0	1	2	3	4	5	6	7	8	9	10	11
1.000	0.462	0.516	0.432	0.444	0.326	0.442	0.288	0.283	0.265	0.202	0.168
12	13	14	15	16	17	18	19	20	21	22	23
0.142	0.100	0.122	0.123	0.085	0.045	0.026	-0.013	0.001	-0.029	-0.049	-0.077
24	25	26	27	28	29	30	31	32			
-0.038	-0.100	-0.095	-0.055	-0.081	-0.092	-0.034	-0.063	-0.089			

High correlations and significant even after 32 months!

ACF – Example: U.S. Stock Dividends

Example (continuation): Correlogram for **US Monthly Changes in Dividends** (1871 – 2020)



Note: Correlations are positive for almost 1.5 years, then become negative.

ACF – Joint Significance Tests

• The application of the LB test to the ACF is straightforward.

Recall that we use the the Ljung-Box (LB) statistic to test H_0: $\rho_1 = \rho_2 = ... = \rho_m = 0$. Under H_0,

$$LB = T(T+2)\sum_{k=1}^{m}\left(\frac{\hat{\rho}_k^2}{(T-k)}\right) \xrightarrow{d} \chi_m^2$$

Example: LB test with **20 lags** for **US Monthly Returns** and **Changes in Dividends** (1871 – 2020)

```
> Box.test(lr_p, lag=20, type= "Ljung-Box")
data:  lr_p
X-squared = 208.02, df = 20, p-value < 2.2e-16
```
⇒ Reject H_0 at 5% level.

```
> Box.test(lr_d, lag=20, type= "Ljung-Box")
data:  lr_d
X-squared = 2762.7, df = 20, p-value < 2.2e-16
```
⇒ Reject H_0 at 5% level.

Conclusion: We found joint significance of first 20 autocorrelations.

Partial ACF (PACF)

• The ACF gives us a lot of information about the order of the dependence when the series we analyze follows a MA process: The ACF is zero after q lags for an MA(q) process.

• If the series we analyze, however, follows an ARMA or AR, the ACF alone tells us little about the orders of dependence: We only observe an exponential decay.

• We introduce a new function that behaves like the ACF of MA models, but for AR models, namely, the partial autocorrelation function (PACF).

• The PACF is similar to the ACF. It measures correlation between observations that are k time periods apart, after controlling for correlations at intermediate lags.

Partial ACF

Intuition: Suppose we have an AR(1):

$$y_t = \phi_1 y_{t-1} + \varepsilon_t.$$

Then,

$$\gamma(2) = \phi_1^2 \gamma(0)$$

The correlation between y_t and y_{t-2} is not zero, as it would be for an MA(1), because y_t is dependent on y_{t-2} through y_{t-1}.

Suppose we break this chain of dependence by removing ("partialing out") the effect y_{t-1}. Then, we consider the correlation between $[y_t - \phi_1 y_{t-1}]$ & $[y_{t-2} - \phi_1 y_{t-1}]$ –i.e, the correlation between y_t & y_{t-2} with the linear dependence of each on y_{t-1} removed:

$\gamma(2)$ = Cov($y_t - \phi_1 y_{t-1}$, $y_{t-2} - \phi_1 y_{t-1}$) = Cov(ε_t, $y_{t-2} - \phi_1 y_{t-1}$) =0

Similarly,

$\gamma(k)$ = Cov(ε_t, $y_{t-k} - \phi_1 y_{t-1}$) = 0 for all $k > 1$.

Partial ACF

Definition: The **PACF** of a stationary time series $\{y_t\}$ is

ϕ_{11}= Corr(y_t, y_{t-1}) = ρ(1)

ϕ_{hh} = Corr(y_t – E[$y_t | I_{t-1}$], y_{t-h} – E[$y_{t-h} | I_{t-1}$]) for h = 2, 3, ...

This removes the linear effects of $y_{t-2}, ..., y_{t-h}$.

• The PACF ϕ_{hh} is also the last coefficient in the **best linear prediction** of y_t given $y_{t-1}, y_{t-2}, ..., y_{t-h}$. (⇒ OLS!)

• Estimation by Yule-Walker equation, using sample estimates:

$$\hat{\boldsymbol{\phi}}_h = [\hat{\boldsymbol{R}}]^{-1}\hat{\boldsymbol{\gamma}}(k) \qquad \Rightarrow \text{a recursive system,}$$

where $\boldsymbol{\phi}_h$ = (ϕ_{h1}, ϕ_{h2} , ..., ϕ_{hh}) and $\boldsymbol{R}$ is the (hxh) correlation matrix.

• OLS is used. Also, a recursive algorithm by Durbin-Levinson.

Partial ACF – AR(p)

Example: AR(p) process:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$
$$E[y_t | I_{t-1}] = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-h-1}$$
$$E[y_{t-h} | I_{t-1}] = \mu + \phi_1 y_{t-h-1} + \phi_2 y_{t-h-2} + \ldots + \phi_p y_{t-1}$$

Then, $\phi_{hh} = \phi_h$ if $1 \leq h \leq p$
$= 0$ otherwise

⇒ After the p^{th} *PACF,* all remaining PACF are 0 for AR(p) processes.

- The plot of the PACF is called the **partial correlogram**.

Partial ACF – AR(p=2)

Example: We simulate an AR(2) process:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

```
sim_ar22 <- arima.sim(list(order=c(1,0,0), ar=c(0.5, 0.3)), n=200)        #simulate AR(2) series
plot(sim_ar22, ylab="Simulated Series", main=(expression(AR(2):~~~phi==c(0.5,0.3))))
pacf_ar22 <- pacf(sim_ar22)
```

Print PACF
> pacf_ar2

1	2	3	4	5	6	7	8	9	10	11
0.558	0.286	0.038	0.103	-0.010	0.009	0.111	0.060	-0.021	-0.076	0.016

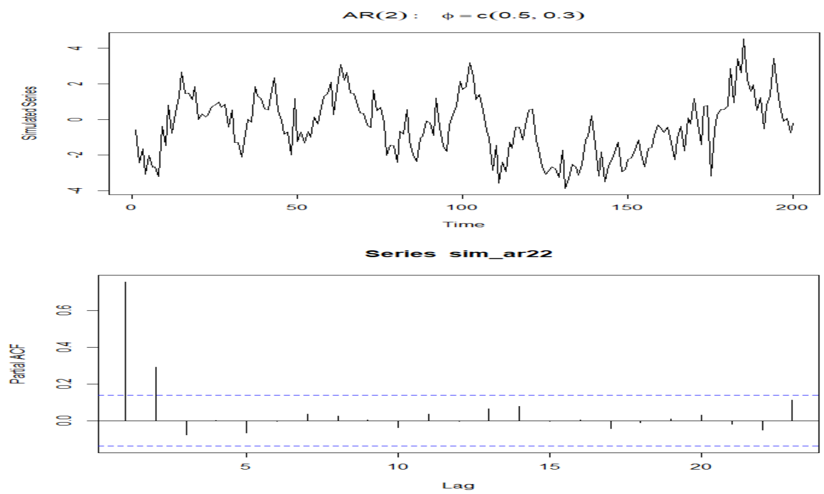
12	13	14	15	16	17	18	19	20	21	22	23
-0.086	-0.139	0.100	0.061	-0.156	0.078	-0.103	0.043	-0.075	0.104	0.024	0.061

SE(r_k) ≈ 1/sqrt(200) = .0707. ⇒ 95% C.I.: ± 2* 0.0707

Partial ACF – AR(p=2)

Example (continuation): Plot of simulated series and PACF

```
> plot(sim_ar22, ylab="Simulated Series", main=(expression(AR(2):~~~phi==c(0.5,0.3))))
> pacf_ar2 <- pacf(sim_ar22)
```



Partial ACF – AR(p=2)

Example (continuation):

Note: The PACF can be calculated by h regressions, each one with h lags. The hh coefficient is the h th order PACF. Using *ar* R function:

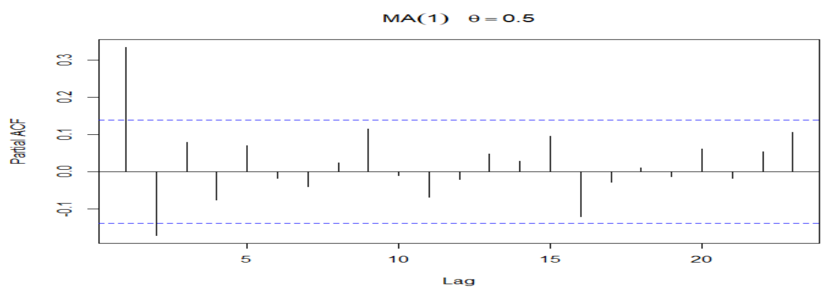
```
> ar(sim_ar2, order.max=1, method = "ols")

Coefficients:
    1
0.5586

Intercept: -0.008403 (0.0761)

Order selected 1  sigma^2 estimated as  1.152


> ar(sim_ar2, order.max=2, method = "ols")

Coefficients:
    1      2
0.3974  0.2869

Intercept: -0.009847 (0.07326)

Order selected 2  sigma^2 estimated as  1.063
```

Partial ACF – MA(q)

• Following the analogy that PACF for AR processes behaves like an ACF for MA processes, we will see exponential decay ("*tails off*") in the partial correlogram for MA process. Similar pattern will also occur for ARMA(p, q) process.

Example: We simulate an MA(1) process with θ_1 = 0.5.

```
sim_ma1 <- arima.sim(list(order=c(0,0,1), ma = 0.5), n=200)
> pacf(sim_ma1)
```

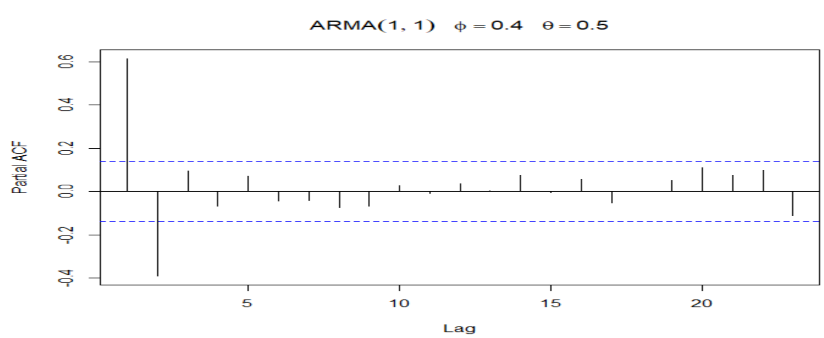


Partial ACF – ARMA(p,q)

• For an ARMA processes, we will see exponential decay ("*tails off*") in the partial correlogram.

Example: We simulate an ARMA(1) process with ϕ_1= 0.4 & θ_1= 0.5.

```
sim_arma11 <- arima.sim(list(order=c(1,0,1), ar=0.4, ma=0.5), n=200)
> pacf(sim_arma11)
```



PACF – Example: U.S. Stock Returns

Example: **US Monthly Returns** (1871 – 2020, T=1,795)

```
pacf_p <- acf(lr_p)                    # pacf: R function that estimates the PACF
> pacf_p
```

Partial autocorrelations of series 'lr_p', by lag

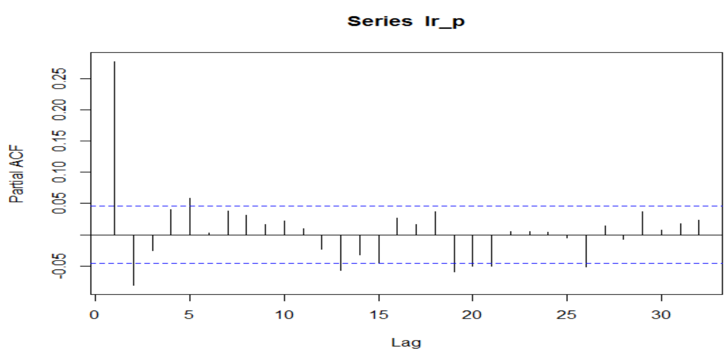
1	2	3	4	5	6	7	8	9	10	11	
0.278	-0.081	-0.026	0.041	0.058	0.002	0.038	0.032	0.016	0.022	0.009	
12	13	14	15	16	17	18	19	20	21	22	23
-0.023	-0.057	-0.032	-0.045	0.027	0.017	0.037	-0.059	-0.051	-0.050	0.005	24
23	24	25	26	27	28	29	30	31	32		
0.006	0.004	-0.005	-0.051	0.014	-0.007	0.037	0.008	0.018	0.023		

$SE(r_k) = 1/\text{sqrt}(1{,}795) = .0236.$ $\Rightarrow$ 95% C.I.: $\pm$ 2* 0.0236

PACF – Example: U.S. Stock Returns

Example (continuation): Correlogram for **US Monthly Returns** (1871 – 2020)

```
> pacf(lr_p)
```



Note: With the exception of the first partial correlation, partial correlations are small, though, again, some are significant.

PACF – Example: U.S. Stock Dividends

Example: **US Monthly Stock Dividends** (1871 – 2020, T=1,795)

```
pacf_d <- pacf(lr_d)
> pacf_d

Partial autocorrelations of series 'lr_d', by lag

    1      2      3      4      5      6      7      8      9     10     11
 0.462  0.385  0.160  0.150 -0.033  0.189 -0.054 -0.056  0.027 -0.082 -0.019
   12     13     14     15     16     17     18     19     20     21     22     23
-0.063 -0.035  0.067  0.043  0.010 -0.057 -0.046 -0.043 -0.008 -0.031 -0.039
    24     25     26     27     28     29     30     31     32
-0.041  0.050 -0.036 -0.030  0.091  0.006 -0.017  0.044 -0.002 -0.042
```

Higher partial correlations than for stock returns.

ARIMA Models: Identification – Correlations

• Correlation approach.

Basic tools: sample ACF and sample PACF.

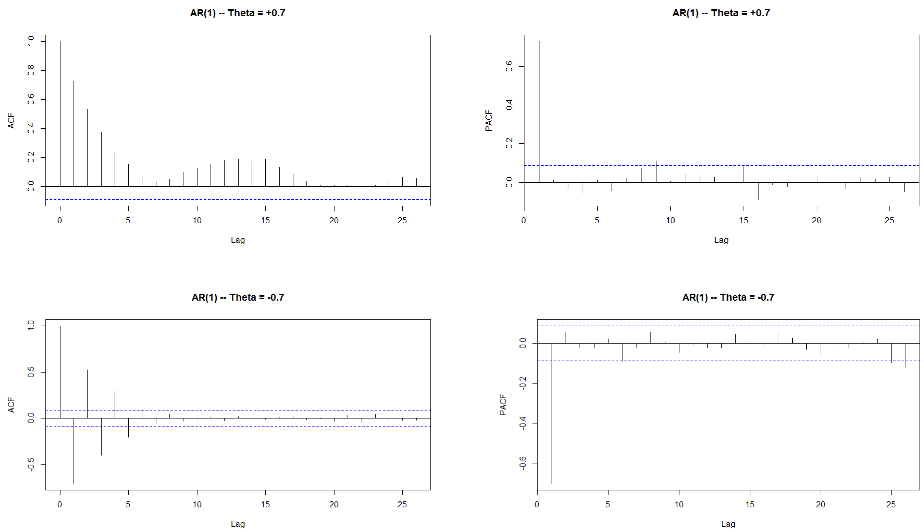
- ACF identifies order of MA: Non-zero at lag q; zero for lags > q.
- PACF identifies order of AR: Non-zero at lag p; zero for lags >p.
- All other cases, try ARMA(p, q) with p > 0 and q > 0.

Summary: For p>0 and q>0.

	AR(p)	MA(q)	ARMA(p, q)
ACF	Tails off	0 after lag q	Tails off
PACF	0 after lag p	Tails off	Tails off

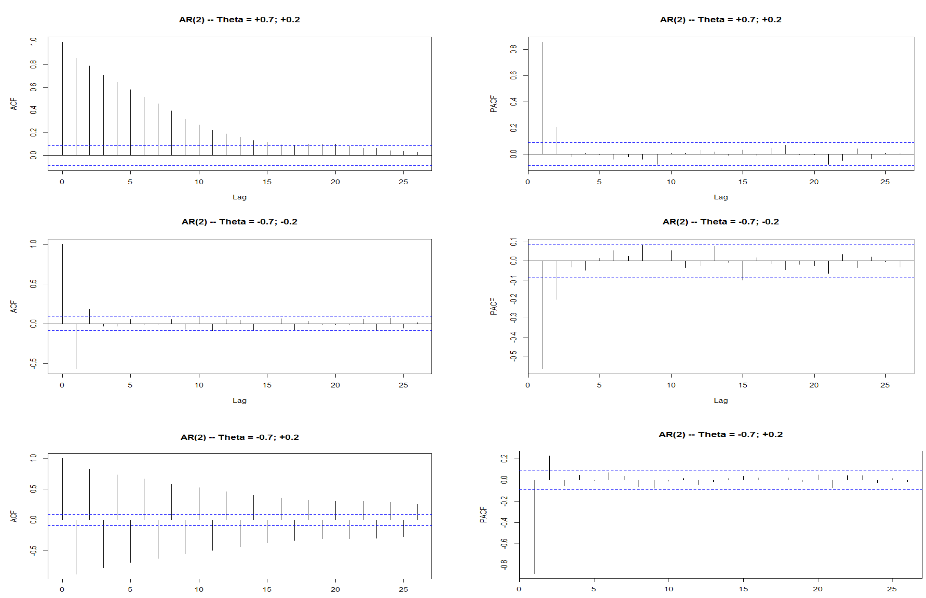
Note: Ideally, "Tails off" is exponential decay. In practice, in these cases, we may see a lot of non-zero values for the ACF and PACF.

ARMA Models: Identification – AR(1)

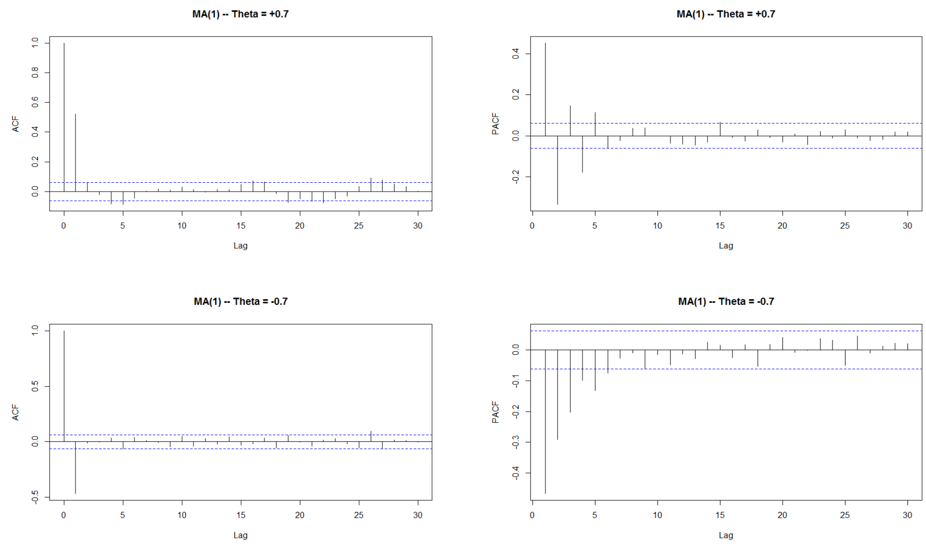


89

ARMA Models: Identification – AR(2)



ARMA Models: Identification – MA(1)



91

ARMA Models: Identification – MA(2)